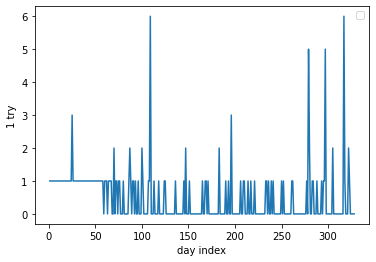
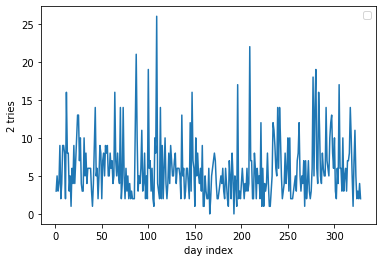
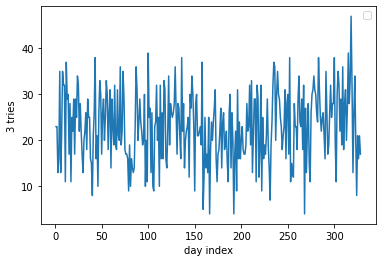
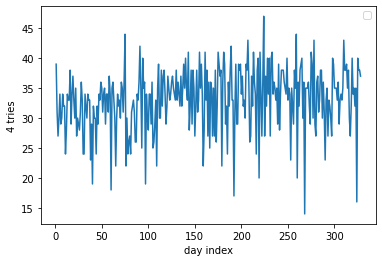
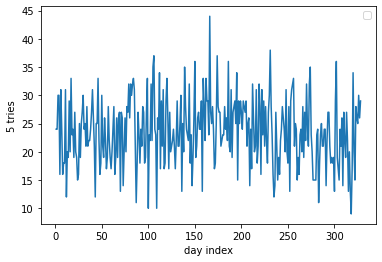
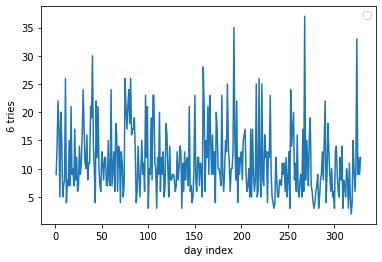
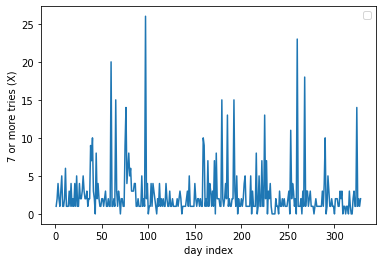
We use the distribution of tries to define difficulty, we cluster them by K-Means, and we also plot the number of tries over time, we don’t see a clear trend of these values over time. Thus, we will use all these values to analyze the difficulty of words.

for and…, more to say here



After we define the difficulty, we use the same model XGboost as question 1 for classification, since we wish to explain the attributes of the word. The loss we use is softmax.

The accuracy of the model on the test set is 73% and 74% of accuracy on the training set, this is model is consistent on the test set and training set. And it classifies “eerie” as hard. Can also check the accuracy of hard class.

